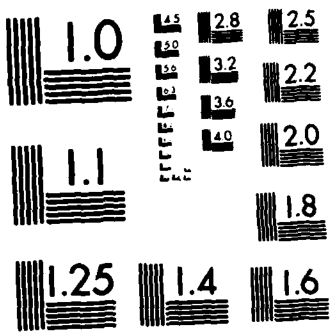
AD-A145 709 ESTIMATION IN NONLINEAR TIME SERIES MODEL II: SOME NONSTATIONARY SERIES(U) NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS D TJOSTHEIM JUL 84 TR-71 1/1

UNCLASSIFIED AFOSR-TR-84-0827 F49620-82-C-0009 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

ESTIMATION IN NONLINEAR TIME SERIES MODELS II:

SOME NONSTATIONARY SERIES

by

Dag Tjøstheim

TECHNICAL REPORT #71

July 1984

ESTIMATION IN NONLINEAR TIME SERIES MODELS II:

SOME NONSTATIONARY SERIES

by

Dag Tjøstheim

Department of Mathematics
University of Bergen
5000 Bergen, Norway

and

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina 27514

Abstract

In an earlier paper (Tjøstheim 1984a) a general framework was introduced for analyzing estimates in stationary nonlinear time series models. In the present paper the framework is enlarged to include certain nonstationary and nonlinear series. General conditions for strong consistency and asymptotic normality are derived both for conditional least squares and maximum likelihood type estimates. Examples are taken from threshold autoregressive, random coefficient autoregressive and doubly stochastic (dynamic state space) models. The emphasis in the examples is on conditional least squares estimates.

Research supported by AFOSR Contract No. F49620 82 C 0009.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S): TR #71

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR-

6a. NAME OF PERFORMING ORGANIZATION: University of North Carolina

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research

6c. ADDRESS (City, State, and ZIP Code): Department of Statistics, Chapel Hill NC 27514

7b. ADDRESS (City, State, and ZIP Code): Directorate of Mathematical & Information Sciences, AFOSR, Bolling AFB DC 20332

8a. NAME OF FUNDING / SPONSORING ORGANIZATION: AFOSR

8b. OFFICE SYMBOL *(If applicable)*: NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620-82-C-0009

8c. ADDRESS (City, State, and ZIP Code): Bolling AFB DC 20332

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2304	A5	

11. TITLE *(Include Security Classification)*: ESTIMATION IN NONLINEAR TIME SERIES MODELS II: SOME NONSTATIONARY SERIES

12. PERSONAL AUTHOR(S): Dag Tjøstheim

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	JUL 84	40

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

In an earlier paper (Tjøstheim 1984a) a general framework was introduced for analyzing estimates in stationary nonlinear time series models. In the present paper the framework is enlarged to include certain nonstationary and nonlinear series. General conditions for strong consistency and asymptotic normality are derived both for conditional least squares and maximum likelihood type estimates. Examples are taken from threshold autoregressive, random coefficient autoregressive and doubly stochastic (dynamic state space) models. The emphasis in the examples is on conditional least squares estimates.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. Brian W. Woodruff

22b. TELEPHONE *(Include Area Code)*: 767-5027

22c. OFFICE SYMBOL: NM

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE: UNCLASSIFIED

1. Introduction

The assumption of stationarity imposed in Tjøstheim (1984a) (hereafter referred to as T1) is sometimes too strict. In this paper we will try to extend the general framework established in T1 to some classes of nonstationary models. We will only treat certain types of nonstationarity, such as that arising from a nonexisting stationary initial distribution, or the nonstationarity arising from a nonhomogeneous generating white noise process $\{e_t\}$. We will also briefly look at autoregressive (AR) models, where the AR coefficients are deterministic functions over time.

The approach to proving consistency and asymptotic normality is similar to the one used in T1. We rely on Theorem 2.1, but need a scaled version of Theorem 2.2 of that paper. The ergodic theorem and the Billingsley (1961) central limit theorem for an ergodic strictly stationary martingale difference sequence will not be available any more, and, as a consequence, a heavier use of pure martingale arguments (mainly martingale type almost sure convergence and central limit theorems) are necessary to obtain our results. The results are not complete, and our examples are not as general as in T1, but we believe that they are representative at least for some of the difficulties arising.

2. Conditional least squares.

Throughout the paper we will use the same notation as in T1. Thus we let $\{X_t, t \in I\}$ be a d-dimensional discrete time stochastic process taking values in R^d and defined on a probability space (Ω, F, P). The second moments of $\{X_t\}$ will be assumed to exist. The index set I is either the set of all integers or the set of all

positive integers. We denote by F_t^X the σ-field generated by $\{X_s,\ s\leq t\}$ and by $\tilde{X}_{t|t-1}(\beta)$ the conditional expectation $E_\beta(X_t|F_{t-1}^X)$, which depends on an r-dimensional parameter vector β. As in T1, we will often suppress β in our notations. Moreover, we denote by $f_{t|t-1}= f_{t|t-1}(\beta)$ the d×d conditional prediction error covariance matrix defined by

$$f_{t|t-1}=E\{(X_t-\tilde{X}_{t|t-1})(X_t-\tilde{X}_{t|t-1})^T|F_{t-1}^X\} \qquad (2.1)$$

We assume that observations $(X_1,..,X_n)$ are given, and we intend to estimate β by minimization of the conditional least squares penalty function given by

$$Q_n(\beta) = \sum_{t=m+1}^{n} |X_t-\tilde{X}_{t|t-1}(\beta)|^2, \qquad (2.2)$$

where $|\cdot|$ is used to denote Euclidean norm, and where in practice the lower summation limit has to be chosen so that $\tilde{X}_{m+1|m}$ is well-defined in terms of the observations $(X_1,...,X_n)$. Unlike the case of consistency for stationary series in T1, it will not be possible to condition on $F_{t-1}^X(m)$, which is the σ-field generated by $\{X_s,\ t-m\leq s\leq t-1\}$. This is because we will rely more on pure martingale arguments, and then we need an increasing sequence of σ-fields. Hence, in this paper we will always condition with respect to F_{t-1}^X. For autoregressive type processes of order p it will then be possible to express $\tilde{X}_{m+1|m}$ in terms of $(X_1,...,X_n)$ if $\min(n,m) \geq p$.

The following two theorems correspond to Theorems 3.1 and 3.2 of T1. We denote by β^0 the true value of β.

<u>Theorem 2.1</u>: Assume that $\{X_t\}$ is a d-dimensional stochastic process with $E\{|X_t|^2\} < \infty$ and such that $\tilde{X}_{t|t-1}(\beta) = E_\beta\{X_t|F_{t-1}^X\}$ is almost surely twice continuously differentiable in an open set B containing β^0. Moreover, assume that there are two positive constants M_1 and M_2 such that for $t\geq m+1$

$$\text{CN1:}\quad E\left\{\frac{\partial \tilde{X}^T_{t|t-1}}{\partial \beta_i}(\beta^0)\ f_{t|t-1}(\beta^0)\ \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}(\beta^0)\right\} \le M_1$$

and

$$\text{CN2:}\quad E\left\{\frac{\partial^2 \tilde{X}^T_{t|t-1}}{\partial \beta_i \partial \beta_j}(\beta^0)\ f_{t|t-1}(\beta^0)\ \frac{\partial^2 \tilde{X}_{t|t-1}}{\partial \beta_i \partial \beta_j}(\beta^0)\right\} \le M_2$$

for $i,j=1,\ldots,r$.

$$\text{CN3:}\quad \liminf_{n\to\infty} \lambda^n_{min}(\beta^0) \overset{a.s.}{>} 0$$

where $\lambda^n_{min}(\beta^0)$ is the smallest eigenvalue of the symmetric non-negative definite matrix $A^n(\beta^0)$ with matrix elements given by

$$A^n_{ij}(\beta^0) = \frac{1}{n}\sum_{t=m+1}^{n} \frac{\partial \tilde{X}^T_{t|t-1}}{\partial \beta_i}(\beta^0)\ \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_j}(\beta^0) \tag{2.3}$$

CN4: Let $N_\delta = \{\beta: |\beta-\beta^0| < \delta\}$ be contained in B. Then

$$\limsup_{n\to\infty}\ \limsup_{\delta\downarrow 0}\ \delta^{-1}\left|A^n_{ij}(\beta) - A^n_{ij}(\beta^0) + \frac{1}{n}\sum_{t=m+1}^{n}\left[\{X_t-\tilde{X}_{t|t-1}(\beta)\}^T\frac{\partial^2 \tilde{X}_{t|t-1}}{\partial \beta_i \partial \beta_j}(\beta) - \{X_t-\tilde{X}_{t|t-1}(\beta^0)\}^T \frac{\partial^2 \tilde{X}_{t|t-1}}{\partial \beta_i \partial \beta_j}(\beta^0)\right]\right| \overset{a.s.}{<} \infty$$

for $i,j=1,\ldots,r$.

Then there exists a sequence of estimators $\hat{\beta}_n = [\hat{\beta}_{n1},\ldots,\hat{\beta}_{nr}]^T$ such that $\hat{\beta}_n \overset{a.s.}{\to} \beta^0$, and such that for $\varepsilon > 0$, there is an event in (Ω,F,P) with $P(E) > 1-\varepsilon$ and an n_0 such that for $n > n_0$, $\partial Q_n(\hat{\beta}_n)/\partial\beta_i = 0$, $i=1,\ldots,r$, and Q_n attains a relative minimum at $\hat{\beta}_n$.

Proof: From the definition of $Q_n(\beta)$ in (2.2) it is easily seen that $\{\partial Q_n(\beta^0)/\partial\beta_i, F^X_n\}$ is a zero-mean martingale. The increments $U_t = \partial Q_t/\partial\beta_i - \partial Q_{t-1}/\partial\beta_i$ are such that (using CN1)

$$E\left(U_t(\beta^0)\right) = 4E\left\{\frac{\partial \tilde{X}^T_{t|t-1}}{\partial \beta_i}(\beta^0) f_{t|t-1}(\beta^0) \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_j}(\beta^0)\right\} \leq 4M_1 \qquad (2.4)$$

and it follows from a martingale strong law of large numbers (cf. Stout 1974, Th. 3.3.8) that $n^{-1}\partial Q_n(\beta^0)/\partial\beta_i \overset{a.s.}{\to} 0$ as $n \to \infty$, and A1 of Theorem 2.1 of T1 is fulfilled. Computing second order derivatives we have

$$\frac{\partial^2 Q_n}{\partial\beta_i \partial\beta_j} = 2\sum_{t=m+1}^{n} \frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta_i} \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_j} - 2\sum_{t=m+1}^{n} \frac{\partial^2 \tilde{X}^T_{t|t-1}}{\partial\beta_i \partial\beta_j}(X_t - \tilde{X}_{t|t-1}). \qquad (2.5)$$

Here $\{\partial^2 \tilde{X}^T_{t|t-1}(\beta^0)/\partial\beta_i\partial\beta_j [X_t - \tilde{X}_{t|t-1}(\beta^0)]\}$ defines a martingale difference sequence with respect to $\{F^X_t\}$ and using CN2 while reasoning as above we have

$$\frac{1}{n}\frac{\partial^2 Q_n}{\partial\beta_i\partial\beta_j}(\beta^0) - \frac{2}{n}\sum_{t=m+1}^{n} \frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta_i}(\beta^0) \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_j}(\beta^0) \overset{a.s.}{\to} 0 \qquad (2.6)$$

as $n\to\infty$, and hence CN3 implies A2 of Theorem 2.1 of T1. Using (2.5) it is seen that CN4 is identical to A3, and the conclusion now follows from Theorem 2.1 of T1. ||

The conditions CN1 and CN2 may be weakened in two directions. According to Theorem 3.3.8 of Stout (1974), M_i, i=1,2, may be replaced with $M_i t^{\alpha_i}$ with $0 \leq \alpha_i < 1$, i=1,2, allowing a moderate growth of moments with t. Using Corollary 3.3.5 of Stout (1974) it is seen that another possibility is to replace CN1 with

$$E\left(\left|\frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta_i}(\beta^0)\{X_t - \tilde{X}_{t|t-1}(\beta^0)\}\right| \left[\log^+\left|\frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta_i}(\beta^0)\{X_t - \tilde{X}_{t|t-1}(\beta^0)\}\right|\right]^{1+\varepsilon}\right) \leq M_1 \qquad (2.7)$$

for i=1,...,r for some $\varepsilon > 0$ and CN2 with the obvious analogue.

When we now turn to the asymptotic distribution of $\hat{\beta}_n$, we cannot rely on Billingsley's (1961) result for ergodic strictly stationary martingale difference sequences which was used in Theorem 3.2 of T1. However, there are more recent results from martingale central limit

theory that can be applied. Typically these require a random scaling factor.

Let $\partial\tilde{X}_{t|t-1}/\partial\beta$ be the d×r matrix having $\partial\tilde{X}_{t|t-1}/\partial\beta_i$, i=1,...,r, as its column vectors and let R_n be the r×r symmetric non-negative definite matrix given by

$$R_n = \sum_{t=m+1}^{n} E\left(\frac{\partial\tilde{X}^T_{t|t-1}}{\partial\beta} f_{t|t-1} \frac{\partial\tilde{X}_{t|t-1}}{\partial\beta}\right) \tag{2.8}$$

Moreover, denote by T_n the stochastic r×r symmetric non-negative definite matrix defined by

$$T_n = \sum_{t=m+1}^{n} \frac{\partial\tilde{X}^T_{t|t-1}}{\partial\beta} \frac{\partial\tilde{X}_{t|t-1}}{\partial\beta} \tag{2.9}$$

We will denote by A^{-I} the Moore-Penrose inverse of a matrix A and by det(A) the determinant of A. Then we have

Theorem 2.2: Assume that the conditions of Theorem 2.1 are fulfilled and assume in addition that

$$\text{DN1:}\quad \liminf_{n\to\infty} n^{-r} \det\{R_n(\beta^0)\} > 0$$

and

$$\text{DN2:}\quad R_n^{-\frac{1}{2}}(\beta^0)\left[\sum_{t=m+1}^{n} \frac{\partial\tilde{X}^T_{t|t-1}}{\partial\beta}(\beta^0)\{X_t-\tilde{X}_{t|t-1}(\beta^0)\}\{X_t-\tilde{X}_{t|t-1}(\beta^0)\}^T \cdot \frac{\partial\tilde{X}_{t|t-1}}{\partial\beta}(\beta^0)\right]R_n^{-\frac{1}{2}}(\beta^0) \xrightarrow{P} I_r$$

where I_r is the identity matrix of dimension r.

Let $\{\hat{\beta}_n\}$ be the estimators obtained in Theorem 2.1. Then

$$R_n^{-\frac{1}{2}}(\beta^0)T_n(\beta^0)(\hat{\beta}_n-\beta^0) \xrightarrow{d} N(0,I_r) \tag{2.10}$$

Proof: Note first of all that $R_n(\beta^0)$ is finite from CN1. Let

$$\frac{1}{2}\frac{\partial Q_n}{\partial\beta} \triangleq S_n = -\sum_{t=m+1}^{n} \frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta}(X_t-\tilde{X}_{t|t-1}) \triangleq \sum_{t=m+1}^{n} \zeta_t \tag{2.11}$$

Since we are dealing with an asymptotic result, as in the proof of Theorem 2.2 of Klimko and Nelson (1978), we may assume that $S_n(\hat{\beta}_n)=0$. Taylor expanding S_n about β^0 and subsequently normalizing with $R_n^{\frac{1}{2}}(\beta^0)$ we have

$$0 = R_n^{-\frac{1}{2}}(\beta^0)S_n(\beta^0) + R_n^{-\frac{1}{2}}(\beta^0)\frac{\partial S_n}{\partial\beta}(\beta_n^*)(\hat{\beta}_n-\beta^0) \tag{2.12}$$

where β_n^* is an intermediate point between $\hat{\beta}_n$ and β^0. Again, reasoning as in the proof of Theorem 2.2 of Klimko and Nelson (1978), in the limit as $n \to \infty$ we may replace β_n^* by β^0. Moreover, using DN1, the boundedness condition CN2 and the orthogonal increment property of a martingale difference sequence, it follows from Chebyshev's inequality that there exists an n_0 such that

$$F_n(\beta^0) \triangleq R_n^{-\frac{1}{2}}(\beta^0) \sum_{t=m+1}^{n}\left[\frac{\partial\zeta_t}{\partial\beta}(\beta^0) - E\{\frac{\partial\zeta_t}{\partial\beta}(\beta^0)\,|\,F^X_{t-1}\}\right] \tag{2.13}$$

is bounded in probability for $n \geq n_0$. Since, from Theorem 2.1, $\hat{\beta}_n \xrightarrow{P} \beta^0$, it follows that $F_n(\beta^0)(\hat{\beta}_n-\beta^0) \xrightarrow{P} 0$, and therefore, when taking distributional limits in (2.12), $R_n^{-\frac{1}{2}}(\beta^0)\partial S_n(\beta_n^*)/\partial\beta$ may be replaced by

$$R_n^{-\frac{1}{2}}(\beta^0)\sum_{t=m+1}^{n} E\left\{\frac{\partial\zeta_t}{\partial\beta}(\beta^0)\,|\,F^X_{t-1}\right\} = R_n^{-\frac{1}{2}}(\beta^0)T_n(\beta^0), \tag{2.14}$$

and hence from (2.10) and (2.12), the theorem will be proved if we can prove that $R_n^{-\frac{1}{2}}(\beta^0)S_n(\beta^0) \xrightarrow{d} N(0,I_r)$.

We use a Cramer-Wold argument. For an r-dimensional vector λ of real numbers it is sufficient to prove that

$$\lambda^T R_n^{-\frac{1}{2}}(\beta^0) S_n(\beta^0) \xrightarrow{d} N(0,\lambda^T\lambda). \tag{2.15}$$

For this purpose we introduce

$$\xi_{tn} = -\lambda^T R_n^{-\frac{1}{2}} \frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta} (X_t - \tilde{X}_{t|t-1}) = \lambda^T R_n^{-\frac{1}{2}} \zeta_t \tag{2.16}$$

Then $\lambda^T R_n^{-\frac{1}{2}} S_n = \sum_{t=m+1}^{n} \xi_{tn}$, and for $\beta=\beta^0$ we have that ξ_{tn}, $m+1\leq t\leq n$, are martingale increments for a zero-mean square integrable martingale array $J_{in} = \sum_{t=m+1}^{i} \xi_{tn}$, $m+1\leq i\leq n$. It is then sufficient to verify the following conditions (cf. Hall and Heyde 1980, Th. 3.2, where the nesting and integrability conditions of that theorem are trivially fulfilled) for $\beta=\beta^0$:

(i) $\max_{m+1\leq t\leq n} |\xi_{tn}| \xrightarrow{P} 0$

(ii) $\sum_{t=m+1}^{n} \xi^2_{tn} \xrightarrow{P} \lambda^T\lambda$

(iii) $E(\max_{m+1\leq t\leq n} \xi^2_{tn})$ is bounded in n.

The condition (ii) follows trivially from the definition of ξ_{tn} and the assumption DN2. Moreover,

$$\max_{m+1\leq t\leq n} \xi^2_{tn} \leq \sum_{t=m+1}^{n} \xi^2_{tn} = \lambda^T R_n^{-\frac{1}{2}} \sum_{t=m+1}^{n} \zeta_t \zeta_t^T R_n^{-\frac{1}{2}} \lambda, \tag{2.17}$$

and using the definition of R_n in (2.8) we have that the expectation of the extreme right hand side of (2.17) is $\lambda^T\lambda$, and (iii) follows from this.

Also, using the technique described in Hall and Heyde (1980, p. 53), for a given $\varepsilon>0$

$$P\,(\max_{m+1\leq t\leq n} |\xi_{tn}| > \varepsilon) = P\,\{\sum_{t=m+1}^{n} \xi^2_{tn} 1(|\xi_{tn}| > \varepsilon) > \varepsilon^2\} \tag{2.18}$$

where $1(\cdot)$ is the indicator function. But

$$\sum_{t=m+1}^{n} E\left\{\xi_{tn}^{2} 1(|\xi_{tn}|>\varepsilon)\right\} = \sum_{t=m+1}^{n} \lambda^{T} R_{n}^{-\frac{1}{2}} E\left\{\zeta_{t}\zeta_{t}^{T} 1(|\lambda^{T} R_{n}^{-\frac{1}{2}}\zeta_{t}\zeta_{t}^{T} R_{n}^{-\frac{1}{2}}\lambda|>\varepsilon)\right\} R_{n}^{-\frac{1}{2}}\lambda, \tag{2.19}$$

and using the definition of ζ_t and the conditions CN1 and DN1 we have that for a given $\delta>0$, there is an n_0 such that for $n>n_0$ and all t, $m+1\le t\le n$,

$$E\left\{\zeta_{t}\zeta_{t}^{T} 1\left(\left|\lambda^{T} R_{n}^{-\frac{1}{2}}\zeta_{t}\zeta_{t}^{T} R_{n}^{-\frac{1}{2}}\lambda\right|>\varepsilon\right)\right\} < \delta \tag{2.20}$$

for $\beta=\beta^0$. Again using CN1 and DN1 there exists an n_1 such that $|R^{-1}_{n,ij}(\beta^0)| \le kn^{-1}$ for $n\ge n_1$; $i,j=1,\ldots r$, and for some constant k. Let $n' = \max(n_0,n_1)$. Then from (2.19) and (2.20) we have for $\beta=\beta^0$ and for $n\ge n'$

$$\sum_{t=n'}^{n} E\{\xi_{tn}^{2}\ 1(|\xi_{tn}| >\varepsilon)\} \le K(\lambda,k)\ \delta, \tag{2.21}$$

where $K(\lambda,k)$ is a constant depending on λ and k but independent of n. On the other hand, using CN1, DN1 and (2.19) it follows at once that for $\beta=\beta^0$

$$\sum_{t=m+1}^{n'} E\{\xi_{tn}^{2} 1(|\xi_{tn}| > \varepsilon)\} \to 0 \tag{2.22}$$

as $n \to \infty$. Using Chebyshev's inequality, (2.21) and (2.22) now implies (i), and the proof is completed. ||

The matrix R_n corresponds to the number of observations in the statement of Theorem 3.2 of T1. In the stationary ergodic case $n^{-1}R_n \to R$ and $n^{-1}T_n \xrightarrow{a.s.} U$ as $n \to \infty$, where U and R are given by (3.6) and condition D1 of T1, and it is seen that (2.10) reduces to (3.18) of T1 then. However, in the nonstationary case we do not require the convergence of $n^{-1}R_n$ and $n^{-1}T_n$, and in fact for the examples to be treated in the next section these quantities do not always converge.

If $T_n - U_n \xrightarrow{a.s.} 0$, where $U_n = E(T_n)$, then the asymptotic covariance matrix of $\hat{\theta}_n$ is given by $U_n^{-1} R_n U_n^{-1}$, which tends to zero by Theorem 2.1.

3. Examples

As in T1 we will illustrate our general results on a variety of nonlinear time series classes. The technical difficulties are larger than in the stationary ergodic case, and, partly to display the essential elements involved more clearly, we will confine ourselves to discussing scalar first order AR type models. Extensions to higher order and vector models will be relatively straightforward in some of the cases. As for the examples in T1 we will generally omit the superscript 0 for the true value of the parameters.

3.1 Threshold autoregressive processes.

These models were originally introduced by Tong (1977) in connection with the analysis of river flow data. The underlying idea is a piecewise linearization of the model by introduction of a local threshold dependence on the amplitude X_t. In the nomenclature of Tong and Lim (1980) a scalar SETAR (m,p,...,p) model is given by

$$X_t - \sum_{i=1}^{p} a_i^j X_{t-i} = e_t^j \tag{3.1}$$

for $[X_{t-1},\dots,X_{t-p}]^T \in F_j, j=1,\dots,m$, where $F_1,\dots,F_m$ are disjoint regions of the p-dimensional Euclidean space R^p, such that $\bigcup_{j=1}^{m} F_j = R^p$. Moreover, $\{e_t^j\}$, j=1,...,m, are independent white noise series consisting of independent identically distributed (iid) variables.

Tong and Lim (1980) consider the numerical evaluation of maximum likelihood estimates of the parameters of the threshold model. In a reply to the discussion of their paper they also mention

the possibility of applying the theory of Klimko and Nelson (1978) to study the properties of these estimates, but we are not aware of any actual work in this direction.

We will only treat the first order AR case (p=1 in (3.1)), and we will assume that there is only <u>one</u> residual process $\{e_t\}$ consisting of zero-mean iid random variables. We can then write (3.1) as

$$X_t - \sum_{j=1}^{m} a^j X_{t-1} H_j(X_{t-1}) = e_t \tag{3.2}$$

where this equation is supposed to hold for $t \geq 2$ with X_1 as an initial variable, and where $H_j(X_{t-1}) = 1(X_{t-1} \in F_j)$, $1(\cdot)$ being the indicator function. There is no explicit time dependence in (3.1) and (3.2). The reason that we did not treat such processes in connection with our study of stationary processes in T1, is that we have not been able to prove the existence of an invariant stationary distribution for the initial variables in the threshold case (cf. Sec. 4.1 of T1). For a general initial variable X_1 it is clear that the process generated by (3.2) will be nonstationary.

<u>Theorem 3.1</u>: Let $\{X_t\}$ be defined by (3.2). Assume that the threshold regions F_j are such that there exist constants $\alpha_j > 0$ so that for all t, $E\{X_t^2 H_j(X_t)\} \geq \alpha_j$, j=1,...,m. Moreover, assume that $|a^j| < 1$, j=1,...,m, $E(X_1^4) < \infty$ and $E(e_t^4) < \infty$. Then there exists a strongly consistent sequence of estimators $\{\hat{a}_n\} = \{[\hat{a}_n^1, \ldots, \hat{a}_n^m]^T\}$ for $a = [a^1, \ldots, a^m]^T$. These estimates are obtained by minimizing the penalty function Q_n of (2.2), and they are jointly asymptotically normal.

<u>Proof</u>: The system of equations $\partial Q_n / \partial a^j = 0, j=1,\ldots,m$, is linear in $a^1, \ldots, a^m$, and it is easily verified that Q_n is minimized by taking

$$\hat{a}_n^j = \frac{\sum_{t=2}^{n} X_t X_{t-1} H_j(X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1})} \tag{3.3}$$

where this exists with probability one since $E\{X_t^2 H_j(X_t)\} \geq \alpha_j$.

Using (3.2) and the independence of the e_t's we have

$$\tilde{X}_{t|t-1} = \sum_{j=1}^{m} a^j X_{t-1} H_j(X_{t-1}) \text{ and } \frac{\partial \tilde{X}_{t|t-1}}{\partial a^j} = X_{t-1} H_j(X_{t-1}), \tag{3.4}$$

while higher order derivatives are zero. Also, it is easily shown that $f_{t|t-1} = E\,(X_t - \tilde{X}_{t|t-1})^2 | F_{t-1}^X\} = E(e_t^2) = \sigma^2$. Since $\partial \tilde{X}_{t|t-1}/\partial a^j$ does not depend on a^k, k=1,...,m, it follows that CN2 and CN4 of Theorem 2.1 are trivially fulfilled. Moreover, using $|a^j| < 1$, $j=1,\ldots,m$, $E(X_1^2) < \infty$ and $E(e_t^2) < \infty$, it follows from (3.2) that $E(X_t^2) \leq K$ for some constant K, and that CN1 of Theorem 2.1 holds.

From the special structure of the derivatives given in (3.4) we have that the matrix A^n in (2.3) in the present case is a diagonal matrix and is given by

$$A^n = \operatorname{diag}\{\frac{1}{n}\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1})\} \tag{3.5}$$

and using the assumption $E\{X_t^2 H_j(X_t)\} \geq \alpha_j$ we have that CN3 of Theorem 2.1 will be fulfilled if we can prove that

$$\frac{1}{n}\sum_{t=1}^{n} X_t^2 H_j(X_t) - \frac{1}{n}\sum_{j=1}^{n} E\{X_t^2 H_j(X_t)\} \overset{a.s.}{\to} 0 \tag{3.6}$$

for j=1,...,m. This will also be the key relationship used in the proof of asymptotic normality.

From the strong law of large numbers we have

$$\frac{1}{n}\sum_{t=1}^{n} e_t^2 - \sigma^2 \overset{a.s.}{\to} 0 \tag{3.7}$$

as $n \to \infty$. Since $H_i(X_t)H_j(X_t) = \delta_{ij}H_i(X_t)$, it follows from (3.2) that

$$e_t^2 = X_t^2 - 2\sum_{j=1}^{m} a^j X_t X_{t-1} H_j(X_{t-1}) + \sum_{j=1}^{m} (a^j)^2 X_{t-1}^2 H_j(X_{t-1}). \tag{3.8}$$

Again inserting from (3.2) we have

$$\sum_{j=1}^{m} a^j X_t X_{t-1} H_j(X_{t-1}) - \sum_{j=1}^{m} (a^j)^2 X_{t-1}^2 H_j(X_{t-1}) = e_t \sum_{j=1}^{m} a^j X_{t-1} H_j(X_{t-1}) \triangleq U_t \tag{3.9}$$

However, it is easily checked that $\{U_t\}$ is a martingale difference sequence with respect to $\{F_t^X\}$, and since $E(U_t^2) = \sigma^2 E\left[\left\{\sum_{j=1}^{m} a^j X_{t-1} H_j(X_{t-1})\right\}^2\right] \le K_1$ for some constant K_1, it follows from the strong law for martingales (Stout, 1974, Th. 3.3.8) that $n^{-1}\sum_{t=2}^{n} U_t \overset{a.s.}{\to} 0$. Inserted in (3.7) and (3.8) this yields

$$\frac{1}{n}\sum_{t=2}^{n} X_t^2 - \frac{1}{n}\sum_{t=2}^{n}\sum_{j=1}^{m} (a^j)^2 X_{t-1}^2 H_j(X_{t-1}) - \sigma^2 \overset{a.s.}{\to} 0 \tag{3.10}$$

Since $E(X_n^2) \le K$ for all n, we have $n^{-1} X_n^2 H_j(X_n) \overset{a.s.}{\to} 0$ as $n \to \infty$ for $j=1,\dots,m$. Furthermore, since $1 = \sum_{j=1}^{m} H_j(X_t)$, an alternative way of writing (3.10) is

$$\sum_{j=1}^{m} \{1-(a^j)^2\} \frac{1}{n}\sum_{t=1}^{n} X_t^2 H_j(X_t) - \sigma^2 \overset{a.s.}{\to} 0. \tag{3.11}$$

On the other hand, since $E(U_t) = 0$ in (3.9), taking expectations in (3.8) and (3.9) and adjusting the summation index as above we have

$$\sum_{j=1}^{m} \{1-(a^j)^2\} \frac{1}{n}\sum_{t=1}^{n} E\{X_t^2 H_j(X_t)\} - \sigma^2 \to 0 \tag{3.12}$$

as $n \to \infty$. Combining (3.11) and (3.12) it follows that

$$\sum_{j=1}^{m} \{1-(a^j)^2\}\left[\frac{1}{n}\sum_{t=1}^{n} X_t^2 H_j(X_t) - \frac{1}{n}\sum_{t=1}^{n} E\{X_t^2 H_j(X_t)\}\right] \triangleq \sum_{j=1}^{m} \{1-(a^j)^2\} Y_{jn} \overset{a.s.}{\to} 0 \tag{3.13}$$

The zero-mean random variables Y_{jn}, $j=1,\dots,m$ are linearly independent

for each fixed n. Since by assumption $|a^j| < 1$ for $j=1,\ldots,m$, the relationship (3.6) follows from (3.13), and this proves the consistency part of the theorem.

Turning now to the proof of asymptotic normality, it is not difficult to verify that the matrix R_n defined in (2.8) in the present case is given by

$$R_n = \sigma^2 \operatorname{diag}\Big[\sum_{t=2}^{n} E\{X_{t-1}^2 H_j(X_{t-1})\}\Big], \tag{3.14}$$

and using the assumption $E\{X_{t-1}^2 H_j(X_{t-1})\} \geq \alpha_j$ for $j=1,\ldots,m$ it follows at once that DN1 of Theorem 2.2 is fulfilled. Moreover, the matrix in DN2 is seen to be given by

$$D_n = \frac{1}{\sigma^2} \operatorname{diag}\left[\frac{\sum_{t=2}^{n} e_t^2 X_{t-1}^2 H_j(X_{t-1})}{E\left\{\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1})\right\}}\right]. \tag{3.15}$$

Since $E(e_t^4) < \infty$ and $|a^j| < 1$, $j=1,\ldots,m$, there exists a K_2 such that $E\{X_{t-1}^4 H_j(X_{t-1})\} \leq K_2$ for all j and t, and thus, using that e_t is independent of F_{t-1}^X, we have $E[\{e_t^2 X_{t-1}^2 H_j(X_{t-1})\}^2] \leq K_2 E(e_t^4)$. From the martingale strong law applied to the martingale difference sequence $\{e_t^2 X_{t-1}^2 H_j(X_{t-1}) - \sigma^2 X_{t-1}^2 H_j(X_{t-1})\}$ it follows that

$$\frac{1}{n}\sum_{t=2}^{n} e_t^2 X_{t-1}^2 H_j(X_{t-1}) - \frac{\sigma^2}{n}\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1}) \xrightarrow{a.s.} 0. \tag{3.16}$$

Using (3.6) and an addition-subtraction argument in (3.15) it follows that $D_n \xrightarrow{a.s.} I_m$ as $n \to \infty$, and thus from Theorem 2.2

$$\operatorname{diag}\left(\frac{\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1})}{\sigma\left[\sum_{t=2}^{n} E\{X_{t-1}^2 H_j(X_{t-1})\}\right]^{\frac{1}{2}}}\right)(\hat{a}_n - a) \xrightarrow{d} N(0, I_m). \| \qquad (3.17)$$

It should be noted that we have asymptotic independence of the estimates $\hat{a}_n^j$, $j=1,\ldots,m$, in the sense that the asymptotic covariance matrix is diagonal. Moreover, taking (3.6) into consideration it is seen that (3.17) may be rewritten as

$$\operatorname{diag}\left(\frac{1}{\sigma}\left[\sum_{t=2}^{n} E\{X_{t-1}^2 H_j(X_{t-1})\}\right]^{\frac{1}{2}}\right) \xrightarrow{d} N(0, I_m) \qquad (3.18)$$

which reduces to the familiar formula $\{nE(X_t^2)\}^{\frac{1}{2}}/\sigma \xrightarrow{d} N(0,1)$ in the ordinary ($m=1$) stationary AR(1) case.

The conditions stated in the theorem can be relaxed. For example it is not necessary to require that the e_t's are identically distributed. It is not difficult to check that the above proof applies to the case where the e_t's are independent and zero-mean, and where $m \le E(e_t^2) \le M$ and $E(e_t^4) \le M'$ for some positive constants m, M and M'. It should also be noted that a similar nonstationary generalization can be made for the exponential autoregressive model treated in Section 4.1 of T1.

The condition $E\{X_t^2 H_j(X_t)\} \ge \alpha_j$ will be satisfied if the regions are chosen so that $P(X_t \in F_j) \ge \gamma_j > 0$ for some positive constants $\gamma_1,\ldots,\gamma_m$, and where $P(X_t=0) \ne P(X_t \in F_{j_0})$ with F_{j_0} being the region containing 0. As an example where such conditions are satisfied consider the case where $\{X_1, e_2, e_3, \ldots\}$ are iid standard normal, where there are only two regions $F_1 = \{x: x \le 0\}$ and $F_2 = \{x: x > 0\}$, and where $a^1 = 0$ and $a^2 = \frac{1}{2}$. Then $P(X_t \ge 0) \ge \frac{1}{2}$, while $P(X_t < 0) \ge \gamma$ for some $\gamma > 0$, since $E(X_t^2)$ is uniformly bounded in t.

3.2 Random coefficient autoregressive processes.

These processes were treated in considerable generality in the stationary case in Sections 4.2 and 6 of T1. Here we will restrict ourselves to a scalar first order model. Extension to higher order vector models involves the same principles, but is notationally more complex.

We assume that $\{X_t\}$ is given on $-\infty<t<\infty$ by

$$X_t - (a+b_t)\, X_{t-1} = e_t \qquad (3.19)$$

where $\{e_t\}$ and $\{b_t\}$ are zero-mean independent processes each consisting of independent variables such that $m_1 \leq E(e_t^2) \leq M_1$ and $E(b_t^2) \leq M_2$, where m_1, M_1 and M_2 are positive constants such that $a^2+M_2<1$. These conditions guarantee that there exists a $F_t^b \vee F_t^e$ - measurable solution of (3.19) with uniformly bounded second moments. This solution can be expressed as

$$X_t = \sum_{i=0}^{\infty} a_{ti} e_{t-i} \qquad (3.20)$$

with $a_{ti} = \prod_{j=0}^{i-1} (a+b_{t-j})$ and where by definition $a_{t0}=1$.

We consider the problem of estimating the parameter a. Since $\tilde{X}_{t|t-1} = aX_{t-1}$, it is clear that there is a unique solution to $\partial Q_n/\partial a = 0$ with Q_n as in (2.2), namely $\hat{a}_n = (\sum_{t=2}^{n} X_t X_{t-1})/(\sum_{t=2}^{n} X_{t-1}^2)$ assuming that observations $(X_1,\ldots,X_n)$ are available. It is our task to find the properties of this estimate.

Theorem 3.2: Let $\{X_t\}$ be as above. If in addition $E(X_t^4) \leq K$ for some constant K, then $\hat{a}_n \to a$.

Proof: It is easily seen that

$$f_{t|t-1} = E\{(X_t - X_{t|t-1})^2 | F_{t-1}^X\} = h_t X_{t-1}^2 + g_t \qquad (3.21)$$

where $h_t = E(b_t^2)$ and $g_t = E(e_t^2)$. Thus

$$E\left\{\left(\frac{\partial \tilde{X}_{t|t-1}}{\partial a}\right)^2 f_{t|t-1}\right\} = E(X_{t-1}^2(h_t X_{t-1}^2 + g_t)) \tag{3.22}$$

is uniformly bounded in t and CN1 of Theorem 2.1 is satisfied. The conditions CN2 and CN4 are trivially satisfied since $\partial \tilde{X}_{t|t-1}/\partial a = X_{t-1}$ is independent of a. It remains to check CN3. This can be done using martingale techniques analogous to those used in the proof of Theorem 3.1.

From (3.21) we have that $\{W_t\} = \{(b_t X_{t-1} + e_t)^2 - (h_t X_{t-1}^2 + g_t)\}$ is a martingale difference sequence with respect to $\{F_t^X\}$. Using our independence assumptions and the fact that uniform boundedness of $E(X_t^4)$ implies uniform boundedness of $E(b_t^4)$ and $E(e_t^4)$ we have

$$E(W_t^2) = E(b_t^4)E(X_{t-1}^4) + 4h_t g_t E(X_{t-1}^2) + E(e_t^4) - h_t^2 E(X_{t-1}^4) - g_t^2 \leq K_1 \tag{3.23}$$

for some constant K_1. From the martingale strong law we have $n^{-1}\sum_{t=2}^{n} W_t \xrightarrow{a.s.} 0$ as $n \to \infty$, and thus, since $b_t X_{t-1} + e_t = X_t - aX_{t-1}$, we have

$$\frac{1}{n}\sum_{t=2}^{n} X_t^2 - \frac{2a}{n}\sum_{t=2}^{n} X_t X_{t-1} + \frac{a^2}{n}\sum_{t=2}^{n} X_{t-1}^2 - \frac{1}{n}\sum_{t=2}^{n} h_t X_{t-1}^2 - \frac{1}{n}\sum_{t=2}^{n} g_t \xrightarrow{a.s.} 0. \tag{3.24}$$

On the other hand, $\{V_t\} = \{(X_t - \tilde{X}_{t|t-1})X_{t-1}\}$ also forms a martingale difference sequence with respect to $\{F_t^X\}$ with $E(V_t^2) \leq C_2$ for some constant C_2 and thus

$$\frac{1}{n}\sum_{t=2}^{n} X_t X_{t-1} - \frac{a}{n}\sum_{t=2}^{n} X_{t-1}^2 \xrightarrow{a.s.} 0. \tag{3.25}$$

Since $E(X_t^2)$ is uniformly bounded, we have $n^{-1}X_n^2 \xrightarrow{a.s.} 0$ as $n \to \infty$, and combining (3.24) and (3.25) it follows that

$$(1-a^2)\frac{1}{n}\sum_{t=2}^{n} X_{t-1}^2 - \frac{1}{n}\sum_{t=2}^{n} h_t X_{t-1}^2 - \frac{1}{n}\sum_{t=2}^{n} g_t \xrightarrow{a.s.} 0$$

Here $h_t \geq 0$ and $|a| < 1$. Hence

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{t=2}^{n} X_{t-1}^2 \geq (1-a^2)^{-1} \liminf_{n \to \infty} \frac{1}{n} \sum_{t=2}^{n} g_t \geq \frac{m_1}{1-a^2} \tag{3.26}$$

and because $\partial \tilde{X}_{t|t-1}/\partial a = X_{t-1}$, this shows that CN3 of Theorem 2.1 is fulfilled and the theorem is proved. ||

It should be noted (cf. Theorem 4.2 of T1) that in the stationary case $E\{X_t^2\} < \infty$ was sufficient to guarantee strong consistency of $\hat{a}_n$. The condition $E(X_t^4) \leq K$ used in the present theorem will be satisfied if $E(e_t^4) \leq C_1$ and $E\{(a + b_t)^4\} \leq C_2 < 1$ for two constants C_1 and C_2. It was needed to obtain a uniform bound on $E(W_t^2)$ in (3.23). Using Corollary 3.3.5 of Stout (1974) it is possible to weaken this to requiring a uniform upper bound on $E\{|W_t|(\log^+|W_t|)^{1+\varepsilon}\}$ for some $\varepsilon > 0$.

As can be expected by analogy with the stationary case treated in T1, boundedness conditions on higher order moments are needed to ensure asymptotic normality.

<u>Theorem 3.3</u>: Let $\{X_t\}$ be as in Theorem 3.2. If in addition $E(e_t^8) \leq C_1$ and $E\{(a+ b_t)^8\} \leq C_2 < 1$ for two constants C_1 and C_2, then $\hat{a}_n$ is asymptotically normal.

<u>Proof</u>: Using (3.22) it is seen that the quantity R_n of (2.8) is given by

$$R_n = \sum_{t=2}^{n} E\{X_{t-1}^2(h_t X_{t-1}^2 + g_t)\}. \tag{3.27}$$

In view of (3.20) we have $g_t E(X_{t-1}^2) \geq g_t E(e_{t-1}^2) = g_t^2 \geq m_1^2$, and it follows that DN1 of Theorem 2.2 is fulfilled.

Employing a subtraction-addition argument and the definition of the quantities used in DN2 it is clear that DN2 will be fulfilled if we can show that

$$\frac{1}{n}\sum_{t=2}^{n} X_{t-1}^2(b_tX_{t-1} + e_t)^2 - \frac{1}{n}\sum_{t=2}^{n} E\{X_{t-1}^2 (h_tX_{t-1}^2 + g_t)\} \xrightarrow{a.s.} 0 \tag{3.28}$$

This will be done in several steps.

First observe that the stated moment conditions on e_t and b_t imply by (3.20) that $E(X_t^8) \leq K_1$ for some constant K_1. Clearly $\{2b_te_tX_{t-1}^3\}$ is a martingale difference sequence with respect to $\{F_t^X\}$, and since $E(X_{t-1}^6) \leq K_2$ for some K_2, it follows from the martingale strong law that

$$\frac{1}{n}\sum_{t=2}^{n} 2b_te_tX_{t-1}^3 \xrightarrow{a.s.} 0. \tag{3.29}$$

Similarly $\{X_{t-1}^2(e_t^2-g_t)\}$ and $\{X_{t-1}^4(b_t^2-h_t)\}$ define martingale difference sequences and using $E(X_t^8) \leq K_1$ and the martingale strong law we have

$$\frac{1}{n}\sum_{t=2}^{n} X_{t-1}^2e_t^2 - \frac{1}{n}\sum_{t=2}^{n} X_{t-1}^2g_t \xrightarrow{a.s.} 0 \tag{3.30}$$

and

$$\frac{1}{n}\sum_{t=2}^{n} X_{t-1}^4b_t^2 - \frac{1}{n}\sum_{t=2}^{n} X_{t-1}^4h_t \xrightarrow{a.s.} 0. \tag{3.31}$$

Inserting (3.29) - (3.31) in (3.28) it is seen that to prove (3.28) it is sufficient to prove

$$\frac{1}{n}\sum_{t=2}^{n} \{X_{t-1}^2 - E(X_{t-1}^2)\}g_t \xrightarrow{a.s.} 0 \tag{3.32}$$

and

$$\frac{1}{n}\sum_{t=2}^{n} \{X_{t-1}^4 - E(X_{t-1}^4)\}h_t \xrightarrow{a.s.} 0 \tag{3.33}$$

Let $Y_t = \{X_t^2 - E(X_t^2)\}u_t$ where $\{u_t\}$ is a positive deterministic sequence bounded above by some constant k and consider the sequence of σ-fields $\{F_t, -\infty < t < \infty\}$, where $F_t = F_t^b \vee F_t^e$. We will prove that $\{Y_t\}$ is a mixingale difference sequence with respect to $\{F_t\}$, i.e.

(cf. Hall and Heyde 1980, p. 19) that for sequences of non-negative constants c_t and ψ_s, where $\psi_s \to 0$ as $s \to \infty$, we have

$$E[\{E(Y_t|F_{t-s})\}^2] \leq \psi_s^2 c_t^2 \tag{3.34i}$$

and

$$E[\{Y_t - E(Y_t|F_{t+s})\}^2] \leq \psi_{s+1}^2 c_t^2 \tag{3.34ii}$$

for all $t \geq 1$ and $s \geq 0$. Since Y_t is F_t-measurable, the condition (ii) in the definition of a mixingale is trivially fulfilled for any choice of c_t and ψ_s.

Using (3.20) and independence properties we have

$$E(X_t^2) = \sum_{i=0}^{\infty} E(a_{ti}^2) g_{t-i} = \sum_{i=0}^{\infty} [\prod_{j=0}^{i-1} E\{(a + b_{t-j})^2\}] g_{t-i} \tag{3.35}$$

and for $s \geq 0$

$$E(X_t^2|F_{t-s}) = \sum_{i=0}^{s-1} [\prod_{j=0}^{i-1} E\{(a + b_{t-j})^2\}] g_{t-i} + \sum_{i=s}^{\infty} \sum_{j=s}^{\infty} E(a_{ti} a_{tj}|F_{t-s}) e_{t-i} e_{t-j} \tag{3.36}$$

However, for $i,j \geq s$ we have

$$E(a_{ti} a_{tj}|F_{t-s}) = \prod_{k=s}^{i-1} (a+b_{t-k}) \prod_{m=s}^{j-1} (a+b_{t-m}) E\{\prod_{k=0}^{s-1} (a+b_{t-k})^2\} \tag{3.37}$$

where by definition $\prod_{k=s}^{i-1} (a+b_{t-k}) = 1$ for $s=i$. Combining (3.35 - 3.37) we obtain

$$|E(Y_t|F_{t-s})| = |u_t E\{\prod_{k=0}^{s-1} (a+b_{t-k})^2\} [\sum_{i=s}^{\infty} \sum_{j=s}^{\infty} \prod_{k=s}^{i-1} (a+b_{t-k}) \prod_{m=s}^{j-1} (a+b_{t-m}) e_{t-i} e_{t-j}$$

$$- \sum_{i=s}^{\infty} E\{\prod_{k=s}^{i-1} (a+b_{t-k})^2\} g_{t-i}]| \leq k(a^2+M_2)^s |\{X_{t-s}^2 - E(X_{t-s}^2)\}| \tag{3.38}$$

and since $E(X_t^4) \leq K$ and $a^2+M_2<1$ it follows that $\{Y_t\}$ is a mixingale difference sequence. Moreover, it follows from the mixingale convergence theorem (Hall and Heyde 1980, Th. 2.21) that (3.32) holds by choosing $u_t = g_{t+1}$.

Next let $Z_t = \{X_t^4 - E(X_t^4)\}u_t$. We will show that $\{Z_t, F_t\}$ defines a mixingale difference sequence. Again the condition (ii) in the definition of a mixingale sequence is trivially fulfilled. Using (3.20) and independence properties we have for $s \geq 0$

$$E(X_t^4|F_{t-s}) = \sum_{i_1=0}^{s-1} \sum_{i_2=0}^{s-1} \sum_{i_3=0}^{s-1} \sum_{i_4=0}^{s-1} E(a_{ti_1}a_{ti_2}a_{ti_3}a_{ti_4}e_{t-i_1}e_{t-i_2}e_{t-i_3}e_{t-i_4})$$

$$+\sum_{i_1=s}^{\infty} \sum_{i_2=s}^{\infty} \sum_{i_3=s}^{\infty} \sum_{i_4=s}^{\infty} E(a_{ti_1}a_{ti_2}a_{ti_3}a_{ti_4}|F_{t-s})e_{t-i_1}e_{t-i_2}e_{t-i_3}e_{t-i_4}$$

$$+ 3 \sum_{i_1=0}^{s-1} \sum_{i_3=s}^{\infty} \sum_{i_4=s}^{\infty} g_{t-i_1} E(a_{ti_1}^2 a_{ti_3}a_{ti_4}|F_{t-s})e_{t-i_3}e_{t-i_4}$$

$$+ 3 \sum_{i_1=0}^{s-1} \sum_{i_4=s}^{\infty} E(e_{t-i_1}^3)\ E\,(a_{ti_1}^3 a_{ti_4}|F_{t-s})e_{t-i_4}). \tag{3.39}$$

A corresponding splitting up of $E(X_t^4)$ yields

$$E(X_t^4) = \sum_{i_1=0}^{s-1} \sum_{i_2=0}^{s-1} \sum_{i_3=0}^{s-1} \sum_{i_4=0}^{s-1} E(a_{ti_1}a_{ti_2}a_{ti_3}a_{ti_4}e_{t-i_1}e_{t-i_2}e_{t-i_3}e_{t-i_4})$$

$$+\sum_{i_1=s}^{\infty} \sum_{i_2=s}^{\infty} \sum_{i_3=s}^{\infty} \sum_{i_4=s}^{\infty} E(a_{ti_1}a_{ti_2}a_{ti_3}a_{ti_4}e_{t-i_1}e_{t-i_2}e_{t-i_3}e_{t-i_4})$$

$$+ 3\sum_{i_1=0}^{s-1} \sum_{i_4=s}^{\infty} g_{t-i_1}g_{t-i_4}E(a_{ti_1}^2 a_{ti_4}^2). \tag{3.40}$$

Since for $i_1, i_2, i_3, i_4 \geq s$, we have that

$$E(a_{ti_1}a_{ti_2}a_{ti_3}a_{ti_4}|F_{t-s}) = \prod_{r=1}^{4}\left\{\prod_{j_r=s}^{i_r-1}(a+b_{t-j_r})\right\} E\left\{\prod_{j=0}^{s-1}(a+b_{t-j})^4\right\} \quad (3.41)$$

it follows that

$$\sum_{i_1=s}^{\infty}\sum_{i_2=s}^{\infty}\sum_{i_3=s}^{\infty}\sum_{i_4=s}^{\infty} E(a_{ti_1}a_{ti_2}a_{ti_3}a_{ti_4}|F_{t-s})e_{t-i_1}e_{t-i_2}e_{t-i_3}e_{t-i_4}$$

$$= E\left\{\prod_{j=0}^{s-1}(a+b_{t-j})^4\right\}X_{t-s}^4 \leq C_2^{s/2}\, X_{t-s}^4 \quad (3.42)$$

Similarly for $i_1 \leq s-1$ and i_3, $i_4 \geq s$ we have

$$E(a_{ti_1}^2 a_{ti_3}a_{ti_4}|F_{t-s}) = \prod_{j=s}^{i_3-1}(a+b_{t-j})\prod_{j=s}^{i_4-1}(a+b_{t-j})E\left\{\prod_{j=0}^{i_1-1}(a+b_{t-j})^4\prod_{j=i_1}^{s-1}(a+b_{t-j})^2\right\} \quad (3.43)$$

and

$$E(a_{ti_1}^3 a_{ti_4}|F_{t-s}) = \prod_{j=s}^{i_4-1}(a+b_{t-j})E\left\{\prod_{j=0}^{i_1-1}(a+b_{t-j})^4\prod_{j=i_1}^{s-1}(a+b_{t-j})\right\} \quad (3.44)$$

It follows that

$$\sum_{i_1=0}^{s-1}\sum_{i_3=s}^{\infty}\sum_{i_4=s}^{\infty} g_{t-i_1}E(a_{ti_1}^2 a_{ti_3}a_{ti_4}|F_{t-s})e_{t-i_3}e_{t-i_4}$$

$$= \sum_{i_1=0}^{s-1} g_{t-i_1}E\left\{\prod_{j=0}^{i_1=1}(a+b_{t-j})^4\prod_{j=i_1}^{s-1}(a+b_{t-j})^2\right\}X_{t-s}^2 \leq M_1 sC_2^{s/4}X_{t-s}^2 \quad (3.45)$$

Moreover, since $|E(e_t^3)| \leq M_3$ for some constant M_3,

$$\left|\sum_{i_1=0}^{s-1}\sum_{i_4=s}^{\infty} E(e_{t-i_1}^3)\; E\,(a_{ti_1}^3 a_{ti_4}|F_{t-s})e_{t-i_4}\right|$$

$$= \left|\sum_{i_1=0}^{s-1}E(e_{t-i_1}^3)\; E\left\{\prod_{j=0}^{i_1=1}(a+b_{t-j})^4\prod_{j=i_1}^{s-1}(a+b_{t-j})\right\}X_{t-s}\right| \leq M_3 sC_2^{s/8}|X_{t-s}| \quad (3.46)$$

Reasoning in an entirely analagous way for the two last terms on the right hand side of (3.40) and inserting in (3.39) and (3.40) we have

$$|E(Z_t|F_{t-s}| \leq u_t[C_2^{s/2}\{X_{t-s}^4 + E(X_{t-s}^4)\}$$
$$+ M_1 sC_2^{s/4}\{X_{t-s}^2 + E(X_{t-s}^2)\} + M_3 sC_2^{s/8}|X_{t-s}|] \quad (3.47)$$

Since $u_t \leq k$, $E(X_t^8) \leq K_1$ and $C_2 < 1$, it follows from the mixingale convergence theorem with $u_t = h_{t+1}$ that (3.33) holds, and the theorem is proved. ||

Again it should be noted that in the stationary case (T1, Th. 4.3) $E(X_t^4) < \infty$ is sufficient to guarantee asymptotic normality of $\hat{a}_n$, while in the present case we require $E(X_t^8) \leq K$.

3.3 Doubly stochastic processes.

Random coefficient autoregressive processes are special cases of what we have termed doubly stochastic time series models in Tjøstheim (1983, 1984b). In the simplest first order case these are given by

$$X_t = \theta_t X_{t-1} + e_t \quad (3.48)$$

where $\{a + b_t\}$ of (3.19) now is replaced by a more general stochastic process $\{\theta_t\}$. The process $\{\theta_t\}$ is usually assumed to be independent of $\{e_t\}$ and to be generated by a separate mechanism. Thus $\{\theta_t\}$ could be a Markov chain or it could itself be an AR process. We refer to Tjøstheim (1983, 1984b) for a definition and properties in the general case.

What makes doubly stochastic processes especially interesting, is that in many cases it is possible to construct recursive forecasting algorithms, and for the case where $\{\theta_t\}$ is an ARMA process, there is a close connection with Kalman type dynamical state space models

(cf. Harrison and Stevens 1976, Ledolter 1981 and Tjøstheim 1983). This type of processes has attracted considerable attention lately, and there exist procedures (see e.g. Ledolter 1981) for computation of unknown parameters, but as far as we know there are no results available concerning the properties of these estimates.

We will only consider a very special case, namely the case where $\{\theta_t\}$ is a first order MA process given by

$$\theta_t = a + \varepsilon_t + b\varepsilon_{t-1}, \tag{3.49}$$

where $\{\varepsilon_t\}$ consists of zero-mean iid random variables independent of $\{e_t\}$ and with $E(\varepsilon_t^2) < \infty$. Both $\{e_t\}$ and $\{\varepsilon_t\}$ will be assumed to be defined on $-\infty < t < \infty$. We will only consider the estimation of a, but we believe that even this simple problem gives a good illustration of the increase in difficulties as we move away from random coefficient autoregressive processes.

To be able to construct Kalman-like algorithms for the predictor $\tilde{X}_{t|t-1}$, the process $\{X_t\}$ must be conditional Gaussian and this requires (Tjøstheim 1983) that $\{e_t\}$ and $\{\varepsilon_t\}$ be Gaussian, and that there is an initial variable X_0 such that the conditional distribution of θ_0 given X_0 is Gaussian. This last requirement is achieved here by choosing $X_0 = 0$. Obviously it implies that $\{X_t\}$ is nonstationary.

<u>Theorem 3.4</u>: Let $\{X_t, t\geq 1\}$ be given by (3.48) and (3.49) under the above stated assumptions. Assume that $E(X_t^4) \leq K$ for some constant K, and that the MA parameter b is less than 1/2 in absolute value. Then there exists a sequence of estimators $\{\hat{a}_n\}$ such that $\hat{a}_n \xrightarrow{a.s.} a$ as $n\to\infty$, and such that $\hat{a}_n$ is obtained by minimization of Q_n in (2.2) as described in the conclusion of Theorem 2.1.

Proof: We follow Tjøstheim (1983) and use the notation $m_t = E(b\varepsilon_t | F_t^X)$ and $\gamma_t = E\{b\varepsilon_t - m_t)^2 | F_t^X\}$. Then it is easily shown from (3.48) and (3.49) that under the stated assumptions we have

$$\tilde{X}_{t|t-1} = (a + m_{t-1})X_{t-1}. \tag{3.50}$$

Moreover, it was shown in Tjøstheim (1983) that $\tilde{X}_{t|t-1}$ can be obtained recursively from the relations

$$m_t = \frac{\delta^2 b X_{t-1}(X_t - aX_{t-1} - m_{t-1}X_{t-1})}{\sigma^2 + X_{t-1}^2(\delta^2 + \gamma_{t-1})} \tag{3.51}$$

with

$$\gamma_t = b^2\delta^2 - \frac{X_{t-1}^2\delta^4 b^2}{\sigma^2 + X_{t-1}^2(\delta^2 + \gamma_{t-1})} \tag{3.52}$$

for $t \geq 1$. Here $\delta^2 = E(\varepsilon_t^2)$ and $\sigma^2 = E(e_t^2)$, while $m_0 = E(b\ \varepsilon_t) = 0$ and $\gamma_0 = E(b^2\varepsilon_t^2) = b^2\delta^2$. It follows that the conditional prediction error is given by

$$f_{t|t-1} = E\{(X_t - \tilde{X}_{t|t-1})^2 | F_{t-1}^X\} = (\delta^2 + \gamma_{t-1})X_{t-1}^2 + \sigma^2. \tag{3.53}$$

From (3.50) and (3.51) we have

$$\frac{\partial \tilde{X}_{t|t-1}}{\partial a} = \left(1 + \frac{\partial m_{t-1}}{\partial a}\right) X_{t-1} \tag{3.54}$$

and, since γ_t is independent of a,

$$\frac{\partial\ m_t}{\partial a} = - \frac{\delta^2 b X_{t-1}^2\left(1 + \frac{\partial m_{t-1}}{\partial a}\right)}{\sigma^2 + X_{t-1}^2(\delta^2 + \gamma_{t-1})}, \tag{3.55}$$

while for $k \geq 2$

$$\frac{\partial^k m_t}{\partial a^k} = -\frac{\delta^2 b X_{t-1}^2 \dfrac{\partial^k m_{t-1}}{\partial a^{k-1}}}{\sigma^2 + X_{t-1}^2(\delta^2 + \gamma_{t-1})} = 0 \tag{3.56}$$

due to the initial condition $m_0=0$. (It is also seen directly from (3.51) that m_t depends linearly on a). It follows that CN2 and CN4 of Theorem 2.1 are trivially fulfilled.

Since $\gamma_{t-1} \geq 0$ we have from (3.55) and the summation formula for a finite geometric series that

$$\left|\frac{\partial m_t}{\partial a}\right| \leq |b| \left(1 + \left|\frac{\partial m_{t-1}}{\partial a}\right|\right) \leq |b| \frac{1 - |b|^t}{1 - |b|} , \tag{3.57}$$

and it follows that $|\partial m_t / \partial a|$ is uniformly bounded. Similarly it follows from (3.52) that $\gamma_t \leq 2b^2\delta^2$. Using (3.53) and (3.54) it is now seen that the condition $E(X_t^4) \leq K$ implies that CN1 of Theorem 2.1 is fulfilled.

Since we assume that $|b| < \frac{1}{2}$, we have by (3.57) that $|\partial m_t / \partial a| < |b| / (1-|b|)<1$ and thus $\liminf_{t \to \infty} (1 + \partial m_t / \partial a)^2 > 0$. From (3.54) and the form of CN3 it is clear that to prove that CN3 holds, it will be sufficient to prove that

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} X_t^2 \overset{a.s.}{>} 0. \tag{3.58}$$

Note that with our initial condion $X_0=0$ we have

$$X_t = \sum_{i=0}^{t-1} a_{ti} e_{t-i} \tag{3.59}$$

where

$$a_{ti} = \prod_{j=0}^{i-1} \theta_{t-j} = \prod_{j=0}^{i-1} (a + \varepsilon_{t-j} + b\varepsilon_{t-j-1}) \tag{3.60}$$

with $a_{t0}=1$. It follows at once that $E(X_t^2) > \sigma^2$, and thus (3.58) will be proved if we can prove that

$$\frac{1}{n}\sum_{t=1}^{n} X_t^2 - \frac{1}{n}\sum_{t=1}^{n} E(X_t^2) \overset{a.s.}{\to} 0. \tag{3.61}$$

This will be done by using the mixingale strong law of large numbers.

Let $Y_t = X_t^2 - E(X_t^2)$ and $F_t = F_t^e \vee F_t^\varepsilon$. Then, Y_t is F_t-measurable and condition (3.34ii) in the definition of a mixingale difference sequence is trivially fulfilled. Moreover, we have from (3.59) and independence properties that for $s \geq 0$

$$E(X_t^2) = \sigma^2 \left\{ \sum_{i=0}^{s-1} E(a_{ti}^2) + \sum_{i=s}^{t-1} E(a_{ti}^2) \right\}, \tag{3.62}$$

where, by definition, the first sum is zero for s=0. It is not difficult to show that

$$E(X_t^2|F_{t-s}) = \sigma^2 \sum_{i=0}^{s-1} E(a_{ti}^2) + 2\sum_{i=0}^{t-1} E(a_{ti}a_{ts}e_{t-i}|F_{t-s})e_{t-s}$$

$$+ \sum_{i=s+1}^{t-1} \sum_{j=s+1}^{t-1} E(a_{ti}a_{tj}|F_{t-s})e_{t-i}e_{t-j}. \tag{3.63}$$

For $i < s$ we have

$$E\{a_{ti}a_{ts}e_{t-i}|F_{t-s}\} = E[E\{a_{ti}a_{ts}e_{t-i}|F_{t-i-1}\}|F_{t-s}]. \tag{3.64}$$

But

$$E\left\{a_{ti}a_{ts}e_{t-i}|F_{t-i-1}\right\} = \left(\prod_{j=i+1}^{s-1} \theta_{t-j}\right) E\left\{\left(\prod_{j=0}^{i-1} \theta_{t-j}^2\right)(a+\varepsilon_{t-i}+b\varepsilon_{t-i-1})e_{t-i}|F_{t-i-1}\right\} \overset{a.s.}{=} 0. \tag{3.65}$$

On the other hand for $i \geq s$

$$E\{a_{ti}a_{ts}e_{t-i}|F_{t-s}\} = e_{t-i}\Big(\prod_{j=s}^{i-1}\theta_{t-j}\Big)E\left\{\Big(\prod_{j=0}^{s-2}\theta^2_{t-j}\Big)(a+\varepsilon_{t-s+1}+b\varepsilon_{t-s})^2|F_{t-s}\right\}$$

$$= e_{t-i}\Big(\prod_{j=s}^{i-1}\theta_{t-j}\Big)\Bigg[a^2+2ab\varepsilon_{t-s}+b^2\varepsilon^2_{t-s})E\Big(\prod_{j=0}^{s-2}\theta^2_{t-j}\Big)+2(a+b\varepsilon_{t-s})E\left\{\Big(\prod_{i=0}^{s-2}\theta_{t-j}\Big)\varepsilon_{t-s+1}\right\}$$

$$+ E\left\{\Big(\prod_{j=0}^{s-2}\theta_{t-j}\Big)\varepsilon^2_{t-s+1}\right\}\Bigg] \triangleq e_{t-i}\Big(\prod_{j=s}^{i-1}\theta_{t-j}\Big)K(t,s) \quad , \tag{3.66}$$

and hence

$$\sum_{i=0}^{t-1} E(a_{ti}a_{ts}e_{t-i}|F_{t-s})e_{t-s} = e_{t-s}X_{t-s}K(t,s). \tag{3.67}$$

Using similar arguments it is not difficult to show that

$$\sum_{i=s+1}^{t-1}\sum_{j=s+1}^{t-1} E(a_{ti}a_{tj}|F_{t-s})e_{t-i}e_{t-j} = X^2_{t-s}K(t,s) \tag{3.68}$$

Inserting in (3.62) and (3.63) we have

$$E(Y_t|F_{t-s}) = E(X^2_t|F_{t-s}) - E(X^2_t) = (2e_{t-s}X_{t-s}+X^2_{t-s})K(t,s)$$

$$-\ \sigma^2 E\left\{\Big(\prod_{j=0}^{s-1}\theta^2_{t-j}\Big)\sum_{i=s}^{t-1}\Big(\prod_{j=s}^{i-1}\theta^2_{t-j}\Big)\right\} \quad . \tag{3.69}$$

Since $E(X^2_t) \leq K_1$ for some K_1, it follows from the proof of Theorem 4.2 of Tjøstheim (1983) that there is a positive $g<1$ such that $E\big(\prod_{j=0}^{s-2}\theta^2_{t-j}\big) = O(g^{s-1})$. Since $E(X^4_t) \leq K$ implies the existence of constants K_2 and K_3 such that $E(e^4_t) \leq K_2$ and $E(\varepsilon^4_t) \leq K_3$, it follows from (3.66) and (3.69) that $E[\{E(Y_t|F_{t-s})\}^2] = O(g^{2(s-1)})$. We then have from the mixingale convergence theorem (Hall and Heyde 1980, Th. 2.21) that (3.61) holds and the theorem is proved. ||

The condition DN2 of Theorem 2.2 is not easy to work with for the present example and we have not ventured to prove asymptotic normality.

<u>3.4 Autoregressive models with deterministic time varying coefficients</u>

Autoregressive models with deterministic time varying coefficients have found applications in several areas, in particular in speech recognition (cf. Markel and Gray 1977), and it is of interest to develop a theory of inference for them. To our knowledge such a theory is largely nonexistent. These models are usually classified as <u>linear</u> nonstationary models so in a sense they fall outside the scope of this paper. However, we will show that at least in special cases it is possible to use the theoretical framework developed in this paper to obtain properties of parameter estimates.

We only look at a first order model, although this is not an essential restriction, and we assume that $\{X_t\}$ is given for all t by

$$X_t - a(t,\alpha)X_{t-1} = e_t \tag{3.70}$$

where $\{e_t\}$ consists of zero-mean iid variables with $E(e_t^2) = \sigma^2$, and where $a(\cdot,\alpha)$ is a deterministic function depending on a scalar parameter α. In this subsection we will use the superscript 0 to denote the true value of α.

<u>Theorem 3.5</u>: Let $\{X_t\}$ be given by (3.70), where $a(t,\alpha)$ is three times continuously differentiable in an open set A containing the true value α^0 of α. Assume that $E(e_t^4) < \infty$, and that there exist positive constants m, M and g with $g<1$ such that for all t

$$|a(t,\alpha^0)| \le g \quad \text{and} \quad \left|\frac{\partial a(t,\alpha^0)}{\partial\alpha}\right| \ge m \tag{3.71}$$

and such that

$$\left|\frac{\partial a^i(t,\alpha)}{\partial \alpha^i}\right| \leq M \tag{3.72}$$

for all t, $\alpha \in A$ and $i=0,1,2$ and 3.

Then there exists a sequence of estimators $\{\hat{\alpha}_n\}$ such that $\hat{\alpha}_n \xrightarrow{a.s.} \alpha^0$, and such that $\hat{\alpha}_n$ is obtained by minimization of Q_n in (2.2) as described in the conclusion of Theorem 2.1. Moreover, $\hat{\alpha}_n$ is asymptotically normal.

<u>Proof:</u> Using (3.72) we can express $\{X_t\}$ as a mean square and almost sure convergent expansion

$$X_t = \sum_{i=0}^{\infty} a_{ti} e_{t-i} \tag{3.73}$$

with $a_{t0}=1$ and $a_{ti} = \prod_{j=0}^{i-1} a(t-j,\alpha)$ for $i > 0$. Moreover, it follows from (3.71) and the mutual independence of the e_t's that

$$E(X_t^2) = \sigma^2 \sum_{i=0}^{\infty} |a_{ti}(\alpha^0)|^2 \leq \frac{\sigma^2}{1-g} . \tag{3.74}$$

From (3.70) and (3.73) we have that

$$\tilde{X}_{t|t-1} = a(t,\alpha)X_{t-1} \quad \text{and} \quad f_{t|t-1} = \sigma^2 \tag{3.75}$$

and since $\partial^i\tilde{X}_{t|t-1}/\partial\alpha^i = \partial a^i(t,\alpha)/\partial\alpha^i \cdot X_{t-1}$, it follows from (3.72), (3.74) and (3.75) that CN1 and CN2 of Theorem 2.1 are fulfilled.

Considering the expression in CN4 of Theorem 2.1, it is seen that by the mean value theorem and (3.72) it is sufficient to prove

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} X_t^2 \overset{a.s.}{<} \infty \tag{3.76}$$

in order to have CN4 satisfied. On the other hand from the equality part of (3.74) we have $E(X_t^2) \geq \sigma^2$, and using (3.71) and the inequality part of (3.74) it follows that both CN3 and CN4 of Theorem 2.1 will be fulfilled if we can prove

$$\frac{1}{n}\sum_{t=1}^{n} X_t^2 - \frac{1}{n}\sum_{t=1}^{n} E(X_t^2) \overset{a.s.}{\to} 0. \tag{3.77}$$

But for $s \geq 0$ we can use (3.73) and (3.74) to show that for $Y_t = X_t^2 - E(X_t^2)$ we have

$$E(Y_t | F_{t-s}^e) = \left\{\prod_{i=0}^{s-1} a^2(t-i,\alpha^0)\right\}\{X_{t-s}^2 - E(X_{t-s}^2)\} \quad . \tag{3.78}$$

It now follows from (3.71), (3.73) and $E(e_t^4) < \infty$, that there is a K such that $E(X_t^4) \leq K$ for all t. Hence, by (3.71) and (3.78) we have $E[\{E(Y_t|F_{t-s}^e)\}^2] \leq g^{2s}K_1$ for some K_1, and the mixingale convergence theorem implies (3.77). The consistency part of the theorem follows from Theorem 2.1.

The quantity R_n defined in (2.8) and used in the proof of asymptotic normality is given in our case by

$$R_n = \sigma^2 \sum_{t=2}^{n} \left|\frac{\partial a(t,\alpha^0)}{\partial\alpha}\right|^2 E(X_{t-1}^2) \quad , \tag{3.79}$$

and it follows at once from $E(X_t^2) \geq \sigma^2$ and the last part of (3.71) that DN1 of Theorem 2.2 is fulfilled. To show that DN2 of that theorem holds it is clearly sufficient to show that

$$\frac{1}{n}\sum_{t=2}^{n}\left|\frac{\partial a(t,\alpha^0)}{\partial\alpha}\right|^2 e_t^2 X_{t-1}^2 - \frac{1}{n}\sigma^2 \sum_{t=2}^{n}\left|\frac{\partial a(t,\alpha^0)}{\partial\alpha}\right|^2 E(X_{t-1}^2) \overset{a.s.}{\to} 0 \tag{3.80}$$

We let $Z_t = |\partial a(t,\alpha^0)/\partial\alpha|^2\{e_t^2X_{t-1}^2 - \sigma^2E(X_{t-1}^2)\}$. Then using (3.73)

and (3.74) and independence properties of $\{e_t\}$ it is not difficult to show that

$$E(Z_t|F^e_{t-s}) = \left|\frac{\partial a(t,\alpha^0)}{\partial\alpha}\right|^2 \sigma^2 \prod_{i=0}^{s-2} a^2(t-1-i,\alpha^0)\{X^2_{t-s} - E(X^2_{t-s})\}. \quad (3.81)$$

From (3.71), (3.72) and $E(X_t^4) \leq K$ it follows that $E[\{E(Z_t|F^e_{t-s})\}^2] \leq g^{2s}M\sigma^2K_1$ for some K_1, and the mixingale convergence theorem implies (3.80). The proof is now concluded by applying Theorem 2.2. ||

It should be realized that the conditions (3.71) and (3.72) are quite restrictive. Thus it is not completely nontrivial to find explicit examples of functions satisfying these requirements.

3.5 Other models

We could of course consider nonstationary versions of bilinear models. But we still face the same obstacles as in the stationary case (cf. Section 4.3 of T1), and again it seems that more progress has to be made on the problem of invertibility before serious analysis of estimates can be undertaken in the present framework.

For the model studied by Aase (1983), however, our theory is applicable and the conditions CN1-CN4 and DN1-DN2 results in conditions which are similar to his, although he considers some slightly different estimates.

4. A maximum likelihood type penalty function.

The maximum likelihood type penalty function was studied in Section 5.1 of T1 and is given in the multivariate case as

$$L_n = \sum_{t=m+1}^{n} [\ln\{\det(f_{t|t-1})\} + (X_t - \tilde{X}_{t|t-1})^T f^{-1}_{t|t-1}(X_t - \tilde{X}_{t|t-1})] \stackrel{\Delta}{=} \sum_{t=m+1}^{n} \phi_t \quad (4.1)$$

The resemblance to a Gaussian log likelihood and the martingale property of L_n was discussed in T1. If $X_t - \tilde{X}_{t|t-1}$ is not independent of F^X_{t-1}, minimization of L_n will in general produce estimates with different properties from those obtained using conditional least squares with Q_n as in (2.2). This is the case for doubly stochastic time series models and in particular for random coefficient autoregressive series.

Corresponding to Theorem 5.1 of T1 and Theorem 2.1 of the present paper we have

Theorem 4.1: Assume that $\{X_t, t\in I\}$ is a d-dimensional process with $E\{|X_t|^2\} < \infty$ for $t\in I$ and such that $\tilde{X}_{t|t-1}(\beta)$ and $f_{t|t-1}(\beta)$ are almost surely twice continuously differentiable in an open set B containing β^0. Moreover, assume that there are positive constants M_1 and M_2 such that for $t \geq m+1$

$$\text{EN1:}\quad E\left\{\left|\frac{\partial\phi_t}{\partial\beta_i}(\beta^0)\right|\right\} \leq M_1$$

and

$$\text{EN2:}\quad E\left[\left|\frac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j}(\beta^0) - E\left\{\frac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j}(\beta^0)\,|\,F^X_{t-1}\right\}\right|^2\right] \leq M_2$$

for i,j=1,...,s, and where expressions for these derivatives are given in (5.8) and (5.9) of T1.

Furthermore, we assume

$$\text{EN3:}\quad \liminf_{n\to\infty} \lambda^n_{min}(\beta^0) \overset{a.s.}{>} 0$$

where $\lambda^n_{min}(\beta^0)$ is the smallest eigenvalue of the symmetric

matrix $C^n(\beta^0)$ with matrix elements given by

$$C^n_{ij}(\beta^0)=\frac{1}{n}\sum_{t=m+1}^{n}\left[\mathrm{Tr}\left\{f^{-1}_{t|t-1}(\beta^0)\frac{\partial f_{t|t-1}}{\partial\beta_i}(\beta^0)f^{-1}_{t|t-1}(\beta^0)\frac{\partial f_{t|t-1}}{\partial\beta_j}(\beta^0)\right\}\right.$$

$$\left.+\ 2\frac{\partial\tilde{X}^T_{t|t-1}}{\partial\beta_i}(\beta^0)f^{-1}_{t|t-1}(\beta^0)\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j}(\beta^0)\right] \tag{4.2}$$

EN4: Let $N_\delta = \{\beta: |\beta-\beta^0| < \delta\}$ be contained in B. Then

$$\lim_{n\to\infty}\sup_{\delta\downarrow 0}(n\delta)^{-1}\left|\sum_{t=m+1}^{n}\left\{\frac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j}(\beta)-\frac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j}(\beta^0)\right\}\right| \overset{a.s.}{<} \infty$$

for $i,j = 1,\ldots,s$.

Then there exists a sequence of estimators $\{\hat{\beta}_n\}$ minimizing L_n of (4.1) such that the conclusion of Theorem 2.1 holds.

Proof: As in the proof of Theorem 2.1 our proof consists in referring our stated conditions back to the conditions of Theorem 2.1 of T1.

From Proposition 5.1 of T1 we have that $\{\frac{\partial\phi_t}{\partial\beta_i}, F^X_t\}$ is a martingale difference sequence, and from a version of the martingale strong law (Stout 1974, Th.3.3.8) it follows from EN1 that $n^{-1}\partial L_n(\beta^0)/\partial\beta_i \overset{a.s.}{\to} 0$ as $n \to \infty$, and A1 of Theorem 2.1 of T1 is fulfilled.

The sequence $\{[\partial^2\phi_t(\beta^0)/\partial\beta_i\partial\beta_j - E\{\partial^2\phi_t(\beta^0)/\partial\beta_i\partial\beta_j|F^X_{t-1}\}], F^X_t\}$ is trivially a martingale difference sequence, and EN2 implies via the just quoted law of large numbers that

$$\frac{1}{n}\left[\frac{\partial^2 L_n}{\partial\beta_i\partial\beta_j}(\beta^0) - \sum_{t=m+1}^{n}E\left\{\frac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j}(\beta^0)\middle|F^X_{t-1}\right\}\right] \overset{a.s.}{\to} 0. \tag{4.3}$$

On the other hand from (5.11) of T1 we have that

$n^{-1}\sum_{t=m+1}^{n} E\{\partial^2\phi_t(\beta^0)/\partial\beta_i\partial\beta_j | F_{t-1}^X\} = C_{ij}^n(\beta^0)$ with $C^n(\beta^0)$ as in (4.2).

It follows from (5.12) of T1 that $C^n(\beta^0)$ is non-negative definite and A2 of Theorem 2.1 of T1 now follows from EN3 and (4.3). Finally, EN4 is just a restatement of A3 of Theorem 2.1 of T1. ||

As for Theorem 2.1 the conditions EN1 and EN2 may be weakened.

We next turn to asymptotic normality and to the analogs of Theorems 5.2 of T1 and 2.2 of the present paper. We let S_n', R_n' and T_n' be the matrices defined by

$$S_n' = S_{n,ij}' = \sum_{t=m+1}^{n} \frac{\partial\phi_t}{\partial\beta_i}\frac{\partial\phi_t}{\partial\beta_j} \ , \tag{4.4}$$

$$T_n' = T_{n,ij}' = E(S_{n,ij}') = \sum_{t=m+1}^{n} E\left(\frac{\partial\phi_t}{\partial\beta_i}\frac{\partial\phi_t}{\partial\beta_j}\right) \tag{4.5}$$

and

$$R_n' = R_{n,ij}' = nC_{ij}^n = \sum_{t=m+1}^{n} E\left\{\frac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j} | F_{t-1}^X\right\} . \tag{4.6}$$

Here expressions for $\partial\phi_t/\partial\beta_i$ and $E(\partial\phi_t/\partial\beta_i \cdot \partial\phi_t/\partial\beta_j)$ are given in (5.8) and (5.18) of T1 and for C^n in (4.2).

Theorem 4.2: Assume that the conditions of Theorem 4.1 are fulfilled and assume in addition that

FN1: $\liminf_{n\to\infty} n^{-s} \det\{R_n'(\beta^0)\} > 0$

and

FN2: $\{R_n'(\beta^0)\}^{-\frac{1}{2}} S_n'(\beta^0)\{R_n'(\beta^0)\}^{-\frac{1}{2}} \xrightarrow{P} I_s$.

Let $\{\hat\beta_n\}$ be the estimators obtained in Theorem 4.1. Then

$$\{R_n'(\beta^0)\}^{-\frac{1}{2}} T_n'(\beta^0)\ (\hat\beta_n - \beta^0) \xrightarrow{d} N(0, I_s).$$

Proof: This is essentially identical to the proof of Theorem 2.1 and is therefore omitted.

5. Two examples.

In Section 6 of T1 it was seen that in the stationary case it was possible to weaken the conditions on the moments of random coefficient autoregressive processes when the maximum likelihood type penalty function was used. The following examples indicate that this continues to hold true for nonstationary doubly stochastic processes. Only consistency will be studied, and the superscript 0 for true values will be dropped.

5.1 A random coefficient autoregressive process.

We will study the first order model given by (3.19), but now we will make the assumption that $E(b_t^2) = \gamma > 0$ is a constant, and we will consider the problem of estimating both a and γ.

Theorem 5.1: Let $\{X_t\}$ be as in (3.19) with $E(b_t^2) = \gamma$. Assume that there exist two positive constants m_1 and M_1 such that $m_1 \leq E(e_t^2) \leq M_1$ and that $a^2 + \gamma < 1$. Then there exists a sequence of estimators $\{[\hat{a}_n, \hat{\gamma}_n]\}$ such that $\{[\hat{a}_n, \hat{\gamma}_n]\} \xrightarrow{a.s.} [a, \gamma]$ and such that $[\hat{a}_n, \hat{\gamma}_n]$ is obtained by minimization of L_n in (4.1) as described in the conclusion of Theorem 2.1.

Proof: For the process treated in this theorem we have that ϕ_t defined in (4.1) is given by

$$\phi_t = \ln(f_{t|t-1}) + (X_t - \tilde{X}_{t|t-1})^2/f_{t|t-1} \tag{5.1}$$

where $\tilde{X}_{t|t-1} = aX_{t-1}$ and $f_{t|t-1} = \gamma X_{t-1}^2 + g_t$ with $g_t = E(e_t^2)$. We have

$$\frac{\partial \phi_t}{\partial a} = -2(X_t - aX_{t-1})/f_{t|t-1} \tag{5.2}$$

and

$$\frac{\partial\phi_t}{\partial\gamma} = \frac{1}{f_{t|t-1}} \frac{\partial f_{t|t-1}}{\partial\gamma} - (X_t - aX_{t-1})^2 \frac{1}{f^2_{t|t-1}} \frac{\partial f_{t|t-1}}{\partial\gamma} \qquad (5.3)$$

Here $1/f_{t|t-1} \leq 1/g_t \leq 1/m_1$, while

$$\frac{1}{f_{t|t-1}} \frac{\partial f_{t|t-1}}{\partial\gamma} = \frac{X^2_{t-1}}{\gamma X^2_{t-1} + g_t} \leq \frac{1}{\gamma} \qquad (5.4)$$

Using similar arguments (cf. also Section 6 of T1) it is not difficult to show that the expectation of the absolute values of the first and second order derivatives of ϕ_t with respect to a and γ are bounded by $K_1 E(X_t^2)$ where K_1 is a constant. However, using independence properties of $\{b_t\}$ and $\{e_t\}$ it is seen from (3.20) that

$$E(X_t^2) = \sum_{i=0}^{\infty} (a^2+\gamma)^i g_{t-i} \begin{cases} \leq (1-a^2-\gamma)^{-1} M_1 \\ \geq (1-a^2-\gamma)^{-1} m_1 \end{cases}, \qquad (5.5)$$

where we have also used $a^2+\gamma<1$. It follows that EN1 and EN2 of Theorem 4.1 are fulfilled.

Since $\gamma>0$ and $a^2+\gamma<1$, there exists an open set B that contains the true parameter vector, and is such that the closure of B in the parameter space do not contain $\gamma = 0$ and $a^2 + \gamma = 1$. Using the martingale law of large numbers it is not difficult to show that there is a constant K_2 such that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{t=2}^{n} \frac{(b_t X_{t-1} + e_t)^2}{\gamma X^2_{t-1} + g_t} \overset{a.s.}{<} K_2 \qquad (5.6)$$

when a and γ are contained in B. Using this result in combination with the above majorizations and with the expression for the third order derivatives (cf. formula (6.8) of T1) we have that EN4 follows from the mean value theorem.

It remains to check EN3. Since $f_{t|t-1}$ does not depend on a, while $\tilde{X}_{t|t-1}$ does not depend on γ, then the matrix C^n of EN3 is given by the diagonal matrix

$$C^n = \frac{1}{n}\begin{bmatrix} \sum_{t=2}^{n} \frac{2X_{t-1}^2}{\gamma X_{t-1}^2 + g_t} & 0 \\ 0 & \sum_{t=2}^{n} \frac{X_{t-1}^4}{(\gamma X_{t-1}^2 + g_t)^2} \end{bmatrix} . \tag{5.7}$$

Assume that there is a subsequence indexed by n_i such that as $n_i \to \infty$, then

$$\frac{1}{n_i}\sum_{t=2}^{n_i} \frac{2X_{t-1}^2}{\gamma X_{t-1}^2 + g_t} \overset{a.s.}{\to} 0. \tag{5.8}$$

Since $2X_{t-1}^2/(\gamma X_{t-1}^2 + g_t) \leq 2\gamma^{-1}$, we can use dominated convergence to show

$$\frac{1}{n_i}\sum_{t=2}^{n_i} E\left(\frac{2X_{t-1}^2}{\gamma X_{t-1}^2 + g_t}\right) \to 0 \tag{5.9}$$

as $n_i \to \infty$. However, from (5.5) we have that $E(X_t^2)$ is bounded uniformly from below, and it follows that there exists a $\delta > 0$ such that $P\{X_t^2 \geq m_1(1-a^2-\gamma)^{-1}\} \geq \delta$, with m_1 as in (5.5). Moreover, the ratio $2x/(\gamma x + g_t)$ is monotonically increasing in $x \geq 0$, so that

$$E\left(\frac{2X_t^2}{\gamma X_{t-1}^2 + g_t}\right) \geq \frac{2m_1(1-a^2-\gamma)^{-1}}{\gamma m_1(1-a^2-\gamma)^{-1} + M_1}\,\delta \tag{5.10}$$

for all t, where M_1 is as in (5.5). But this contradicts (5.9) and we must have

$$\liminf_{n\to\infty} C_{11}^n = \liminf_{n\to\infty} \frac{1}{n}\sum_{t=2}^{n} \frac{2X_{t-1}^2}{\gamma X_{t-1}^2 + g_t} \overset{a.s.}{>} 0$$

It follows using the same argument that $\liminf_{n\to\infty} C_{22}^n \overset{a.s.}{>} 0$, and the

condition EN3 of Theorem 4.1 is verified.||

5.2 A doubly stochastic process

We will only treat the simple example studied in Theorem 3.4.

Theorem 5.2: Let $\{X_t, t \geq 1\}$ be as in Theorem 3.4 with the exception that we replace the condition $E(X_t^4) \leq K$ with the weaker condition $E(X_t^2) \leq K$ for some K. Then there exists a sequence of estimators $\{\hat{a}_n\}$ such that $\hat{a}_n \xrightarrow{a.s.} a$ and such that $\hat{a}_n$ is obtained by minimization of L_n in (4.1) as described in the conclusion of Theorem 2.1.

Proof: Again we have that the functional form of ϕ_t is given by (5.1), but now with $\tilde{X}_{t|t-1}$ and $f_{t|t-1}$ given by (3.50) and (3.53). It follows that

$$\frac{\partial \phi_t}{\partial a} = -2\{X_t - (a+m_{t-1})X_{t-1}\}\left[1 + \frac{\partial m_{t-1}}{\partial a}\right]X_{t-1}/f_{t|t-1} \qquad (5.12)$$

and using the fact that $\partial^k m_{t-1}/\partial a^k = 0$ for $k > 1$ we have

$$\frac{\partial^2 \phi_t}{\partial a^2} = 2\left(1 + \frac{\partial m_{t-1}}{\partial a}\right)^2 X_{t-1}^2/f_{t|t-1} = E\left[\frac{\partial^2 \phi_t}{\partial a^2} \mid F_{t-1}^X\right] \qquad (5.13)$$

while higher order derivatives of ϕ_t are zero. From (3.53) we have $1/f_{t|t-1} \leq 1/\sigma^2$. On the other hand it was proved in the proof of Theorem 3.4 that $|\partial m_{t-1}/\partial a|$ is bounded above by a constant independent of t, and since $E(X_t - \tilde{X}_{t|t-1})^2 \leq E(X_t^2)$, it now follows that $E(|\partial\phi_t/\partial a|) \leq K_1$ for some constant K_1. Thus condition EN1 of Theorem 4.1 holds. Conditions EN2 and EN4 are trivially satisfied and it remains to verify EN3.

The matrix C^n of EN3 in this case reduces to a scalar, namely

$$C^n = \frac{1}{n}\sum_{t=2}^{n} E\left(\frac{\partial^2\phi_t}{\partial a^2}\,\middle|\,F^X_{t-1}\right) = \frac{1}{n}\sum_{t=2}^{n}\left(1+\frac{\partial m_t}{\partial a}\right)^2 \frac{X^2_{t-1}}{(\delta^2+\gamma_{t-1})X^2_{t-1}+\sigma^2} \tag{5.14}$$

It was established in the proof of Theorem 3.4 that

$\liminf_{t\to\infty} (1+\partial m_t/\partial a)^2 \overset{a.s.}{>} 0$ and that $\gamma_t \le 2b^2\delta^2$. Reasoning in the same way as in the last part of the proof of Theorem 5.1 it is concluded that $\liminf_{n\to\infty} C^n \overset{a.s.}{>} 0$, and the theorem is proved. ||

Unfortunately we have not been able to prove asymptotic normality for any of the two examples treated here. The difficulty lies in verifying condition FN2 of Theorem 4.2.

6. Summary remarks

In this paper as well as in T1 we have developed a general framework for analyzing estimates in nonlinear time series models. We have given applications to a number of different model classes and tried to deduce sufficient conditions for strong consistency and asymptotic normality from the general conditions. Our conditions reduce to the standard set of conditions (cf. Fuller 1976, Ch. 8) in the linear case, except that we do not necessarily require a homogeneous residual process $\{e_t\}$.

Explosive behavior, e.g. $E(X_t^2)$ increases as an exponential function of t as $t \to \infty$, is not permitted in the present set up. It should be noted, however, that Lai and Wei (1982, 1983) have recently proved consistency, but not asymptotic normality, of parameter estimates in linear explosive models. It is sometimes difficult to find conditions guaranteeing nonexplosive behavior for nonlinear models, and it is therefore a challenging task to try to extend Lai and Wei's results

to nonlinear series.

Our work has potential applications in several other directions. One would be to extend our results to more general classes of examples, especially in the doubly stochastic case. Another important problem is that of hypothesis testing, in particular in connection with empirical identification of models.

References

Aase, K.K. (1983): Recursive estimation in non-linear time series models of autoregressive type, J. Roy. Stat. Soc. Ser. B, 45, 228-237.

Billingsley, P. (1961): The Lindeberg-Lèvy theorem for martingales, Proc. Amer. Math. Soc., 12, 788-792.

Fuller, W.A. (1976): Introduction to Statistical Time Series, Wiley, New York.

Hall, P. and Heyde, C.C. (1980): Martingale Limit Theory and Its Application, Academic Press, New York.

Harrison, P.J. and Stevens, C.F. (1976): Bayesian forecasting (with discussion), J. Roy. Stat. Soc. Ser. B, 38, 205-248.

Klimko, L.A. and Nelson, D.I. (1978): On conditional least squares estimation for stochastic processes, Ann. Statist., 6, 629-642.

Lai, T.L. and Wei, C.Z. (1982): Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, Ann. Statist. 10, 154-166.

Lai, T.L. and Wei, C.Z. (1983): Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters, J. Multivariate Anal., 13, 1-23.

Ledolter, J. (1981): Recursive estimation and forecasting in ARIMA models with time varying coefficients. In Applied Time Series Analysis II, D.F. Findlay, Ed., Academic Press, New York, 449-471.

Markel, J.D. and Gray, Jr., A.H. (1977): Linear Prediction of Speech, Springer, New York.

Stout, W.F. (1974): Almost Sure Convergence, Academic Press, New York.

Tjøstheim, D. (1983): Some doubly stochastic time series models, Technical Report, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27514, to appear in J. Time Series Anal.

Tjøstheim, D. (1984a): Estimation in nonlinear time series I: Stationary series, Technical Report, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27514.

Tjøstheim, D. (1984b): Recent developments in nonlinear time series modelling, Invited paper, The John B. Thomas Birthday Volume of Papers on Communication, Networks and Signal Processing, Springer, New York, to appear.

Tong, H. (1977): Discussion of a paper by A.J. Lawrance and N.T. Kottegoda, J. Roy. Stat. Soc., Ser. A, 140, 34-35.

Tong, H. and Lim, K.S. (1980): Threshold autoregression, limit cycles and cyclical data (with discussion), J. Roy. Stat. Soc. Ser. B, 42, 245-292.